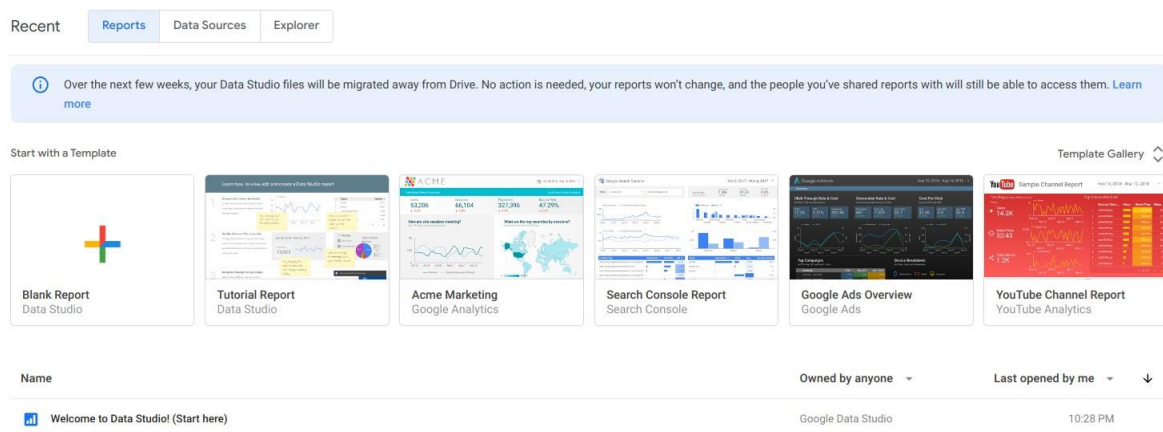# Data Science e Tecnologie per le Basi di Dati Lab 2 – Data Studio

1. Login

Connect to Google Data Studio, login with your Google Account or create a new free Google Account.

- https://datastudio.google.com

2. Welcome report

Click on "Tutorial Report" and follow the tutorial.

- https://datastudio.google.com/reporting/0B5FF6JBKbNJxOWItcWo2SVVVeGc
- learn the basics of the Data Studio tool by copying the "Welcome report" and following the stepby-step instructions provided.
- page 8, "Track report usage with Google Analytics", can be safely skipped.

3. New report on Airbnb Boston reviews

To create a new report from scratch, a data source must be identified. To this aim, a portion of the Kaggle dataset of the Airbnb reviews in Boston has been uploaded into a shared Google Sheets to be used as data source for Google Data Studio.

- the Google Sheets, with approximately 10k reviews to be used as data source, is available at https://docs.google.com/spreadsheets/d/1a2c9vCMFFfDXmhjoEoX2EwS2lYTbqE4WfZY72TXW9co/edit#gid=285360760

- Spend some time to understand the data by reading their description on Kaggle and looking at the table on Google Sheets.
- The data source table has been created by joining the "Listings" and "Reviews" original tables provided by Kaggle, and exporting the first 10k joined rows sorted by ascending "listing_id".

Data sources

Data sources have two types of fields: dimensions and metrics.

- A **dimension** is a category of data.
- A **metric** is a number that quantifies something in that category.
- A Data Studio report lets you visualize those dimensions and metrics in charts and tables.
- In your Data Studio data sources and report properties panels, dimensions appear as **green** chips, while metrics appear as **blue** chips.

Field	ID
Browser	t0._ga:browser_
Sessions	t0._ga:sessions_

Create a new report

- Go to the Data Studio home page.
- Click on "Start a new report" (Blank).

Start with a Template

Blank Report
Data Studio

Tutorial R
Data Studi
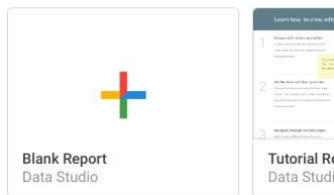
- Close the default window "Add data to report" and rename the "Untitled Report" with a name of your choice by clicking on the name itself.

Untitled Report

File   View   Page   Help

- Create a new data source by clicking on the button "Add data" on the menu at the top.
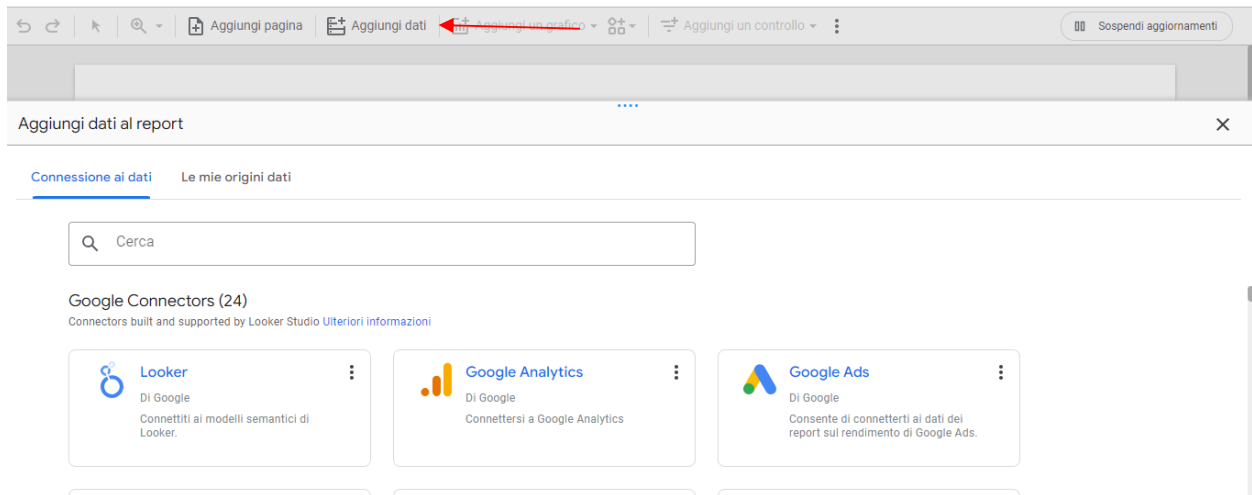
Connect to the Google Sheet data source by using its URL:

- Choose the "Google Sheets" connector from the list of possible connectors
- Choose the "URL" option in the first column
- Paste the Airbnb-data Google Sheet URL in the specific field:
  https://docs.google.com/spreadsheets/d/1a2c9vCMFFfDXmhjoEoX2EwS2lYTbqE4WfZY72TXW9co/edit#gid=285360760
- Choose the "Reviews Query DW" worksheet in the next column
- Tick the option to "use the first row as headers" if it is not ticked yet
- Click on the "Add" button to execute the connection to the data source
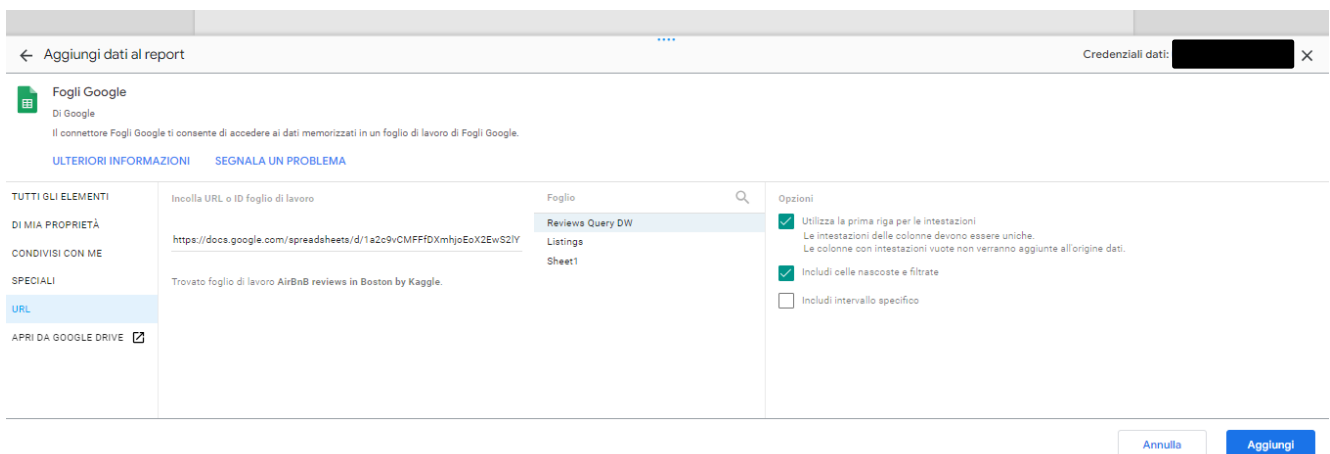
Dimensions, metrics, and transformations

- Check the **type** and **aggregation** of each field and that all the fields are correctly interpreted as either **dimension** or **metric**.
- **Create new useful fields** from the existing ones by exploiting formulas, such as in the following (click on the "Add field" button at the bottom right of the menu). For details on this step, see: https://support.google.com/datastudio/answer/6299685?hl=en
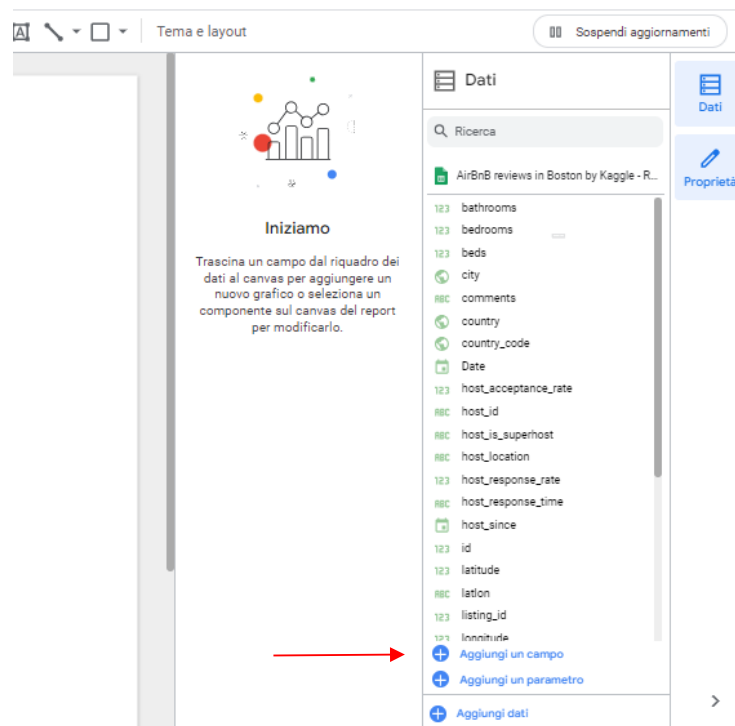
After the definition of a command or formulas, you must click on the button "Done".

A list of examples you can do:

- **LENGTH**(comments) → to count the number of chars of the comment field
- **CONCAT**(latitude, CONCAT(', ', longitude)) → to generate a (lat, long) field useful for map charts; before generating this new field, set "**Type=Text**" for latitude and longitude fields, so that they become dimensions (by default, Data Studio considers them as metrics)
- **price / square_feet** → to compute the average price per square feet (try to create a field that contains the square meters instead of the square feet (1 foot = 0.3048 meter)).
- **MONTH**(Date) → to extract the month of the year from the full date, e.g. 12
- **YEAR**(Date) → to extract the year from the full date, e.g. 2017
- **CONCAT**(YEAR(Date), MONTH(Date)) → to build a field which is the full month, e.g. 201712

■ if you already have the computed fields "month" and "year", you can also use them in the formula, e.g., CONCAT(year, month

← EDIT CONNECTION

Index	Field		Type			Aggregation	
21	property_type	⋮	ABC	Text	▼	None	
22	room_type	⋮	ABC	Text	▼	None	
23	bathrooms	⋮	123	Number	▼	None	▼
24	bedrooms	⋮	123	Number	▼	None	▼
25	beds	⋮	123	Number	▼	None	▼
26	square_feet	⋮	123	Number	▼	None	▼
27	price	⋮	123	Number	▼	None	▼
28	review_scores_rating	⋮	123	Number	▼	None	▼
29	review_scores_value	⋮	123	Number	▼	None	▼
30	comment_length fx	⋮	123	Number	▼	None	▼
31	latlong fx	⋮	🌐	Latitude, Longitude	▼	None	
32	price_per_ft2 fx	⋮	123	Number	▼	None	▼
33	month fx	⋮	📅	Month (MM)	▼	None	
34	year fx	⋮	📅	Year (YYYY)	▼	None	
35	month_year fx	⋮	ABC	Text	▼	None	

↻ REFRESH FIELDS
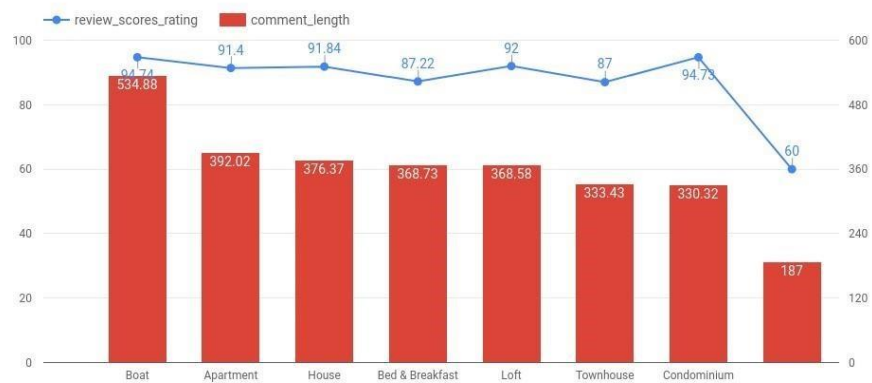
Analyse the data

Analyse the data by building the following visualizations. Then, explore and create new visualizations to find interesting insights on your own.

- **Analysis (1)**: compare the trend of the average length of the review "comments" (number of chars) vs the average "review_scores_rating" for different "propert_type". Sort the data by descending average length of comments. Allow end-users to filter the data under analysis by selecting a date range of their choice.

- **Analysis (2)**: compare the trend of the number of different "listing_id" reviewed, for each "room_type", and for each month of the year. Allow end-users to filter the data under analysis by selecting a date range and the type of superhost (true/false).
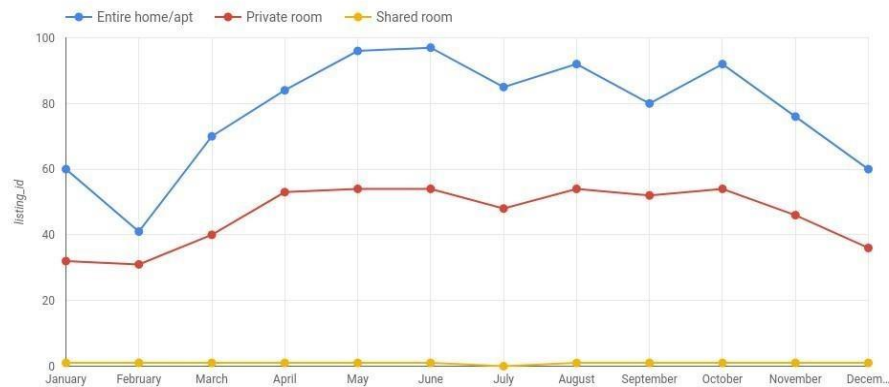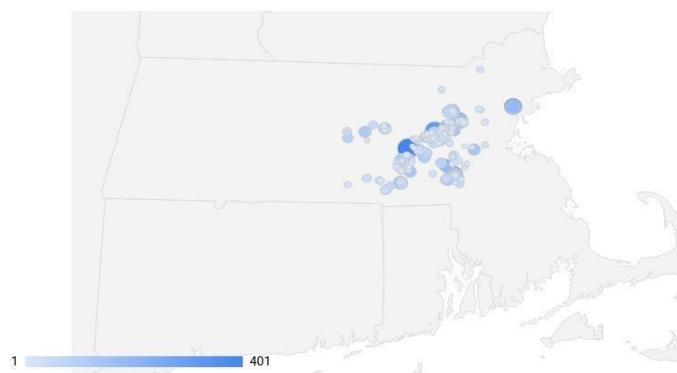

Explore, create and present new additional analyses to identify interesting insights. For instance:

- **Analysis (3)**: analyse the number of different reviewers for each location (lat, long).

   Note that the Kaggle dataset of the Airbnb reviews is in Boston, **Massachusetts**, US

- **Analysis (4)**: Visualize, for each property type and for each year, the average rating score values sorted by ascending property type and by descending mean rating_score_value. Exclude possible null values for the attribute property_type.

Average review score values
for each property type and year

	property_type ❶ ▲	year	review_scores_value ❷ ▼
1.	Apartment	2009	9.38
2.	Apartment	2010	9.04
3.	Apartment	2016	9.03
4.	Apartment	2014	9
5.	Apartment	2013	8.99
6.	Apartment	2015	8.97
7.	Apartment	2011	8.94
8.	Apartment	2012	8.92
9.	Bed & Breakfast	2014	8.94
10.	Bed & Breakfast	2013	8.93

1 - 44 / 44 < >

- **Analysis (5)**: Visualize, for each year and for each room type, the total count of top-scored reviews (review_score_value = 10).

Compare the obtained results with the count of the distinct listing_id reviewed.

Number of reviews
for each room type and year